General instructions. The exam has 2 parts. You have 3 hours. No books, notes, cell phones, or other external materials are allowed.

Part I (40 points)

Answer 4 of the following 5 questions. Give short answers, sketching the proof or giving a convincing justification in 2-5 sentences (even for true/false questions). You may use without proof theorems stated in class, provided you state the appropriate theorem that you are using. As usual, assume $n$ (or $|V|$) is large enough.

a. Suppose $n$ balls are thrown independently and uniformly at random into $n$ bins. Now merge the first $k := 88 \log n$ bins into one new bin, do the same for the next $k$ original bins, and so forth. Assume $n$ is divisible by $k$, hence there are exactly $n/k$ new bins in total.

Is it true that with high probability, all the new bins are non-empty?

b. Does there exist a random variable $X$ taking real values such that $\Pr[X = 4] = \Pr[X = 10] \geq 1/3$ and $\text{Var}(X) = 1$?

c. Let $G = (V, E)$ be an input graph with edge capacities $c : E \rightarrow \mathbb{R}_+$. Suppose we want to find a partition $V = V_1 \cup V_2 \cup V_3$ into 3 equal-size parts $|V_1| = |V_2| = |V_3|$, that has minimum total capacity (defined as $\sum_{i<j} c(V_i, V_j)$).

Show that solving this problem on a $(1 + \varepsilon)$-cut sparsifier $G' = (V, E')$ with edge capacities $c' : E \rightarrow \mathbb{R}_+$, gives a $(1 \pm \varepsilon)$-approximation for this problem on $G$.

d. Consider a bipartite graph on two sets of $n$ nodes, which are connected by $\frac{3}{4} n$ random edges. Is it true that with probability at least $1 - 1/\sqrt{n}$ it will have no cycle?

e. Are there directed graphs on $n$ nodes, where the cover time (the expected time until all nodes have been visited) is $\Theta(2^{\sqrt{n}})$?

Part II (60 points)

Answer 3 of the following 4 questions.
1. Let $B$ be a randomized algorithm that approximates some function $f(x)$ as follows:

$$\forall x, \quad \Pr \left( B(x) \in (1 \pm \varepsilon) f(x) \right) \geq \frac{2}{3}.$$ 

Let algorithm $C$ output the median of $O(\log \frac{1}{\delta})$ independent executions of algorithm $B$ on the same input, for $\delta \in (0, \frac{1}{2})$. Prove that

$$\forall x, \quad \Pr \left[ C(x) \in (1 \pm \varepsilon) f(x) \right] \geq 1 - \delta.$$ 

2. Suppose five players have each a private input in $[n]^n$, denoted respectively $x^j$ for $j \in [5]$. Another player, called the referee, receives from each of these players a short message, denoted respectively $m^j$ for $j \in [5]$, and the referee then finds $\alpha_j \in \{-n, -n + 1, -n + 2, \ldots, n\}$ for $j \in [5]$, that minimize $\| \sum_j \alpha_j x^j \|_2$.

Design a protocol using shared randomness, that approximates this minimization task within factor 2, and analyze its accuracy and message size.

3. Show that it is possible to color the edges of $K_n$, the complete graph on $n$ vertices, with $O(\sqrt{n})$ colors, so that no triangle is monochromatic (meaning that all its edges have the same color).

4. Given $n$ records consisting of student name and gender (Male, Female), Suggest a way of storing the students’ genders, so that later, given a query “student name”, the algorithm can report the student’s correct gender. If the query consists of a name that is not in the list, then any response is acceptable. The goal is to use only $O(n)$ many bits of storage and to allow quick decision on the gender.

Suppose that you have a collection of ideal random functions at your disposal.

(a) Suggest a (randomized) algorithm based on Bloom Filters that errs with probability at most $\epsilon$ for some fixed $\epsilon$. Also suggest one that does not err but may return ‘don’t know’ with probability $\epsilon$.

(b) Suggest an algorithm based on Cuckoo Hashing that does not err at all (whp).

Good Luck.